

iScience, Volume 24

Supplemental information

**The chromosome-scale genome
of *Magnolia officinalis* provides insight
into the evolutionary position of magnoliids**

Yanpeng Yin, Fu Peng, Luoqing Zhou, Xianmei Yin, Junren Chen, Hongjin Zhong, Feixia Hou, Xiaofang Xie, Li Wang, Xiaodong Shi, Bo Ren, Jin Pei, Cheng Peng, and Jihai Gao

Table S1. Raw reads statistics. Related to Table 1

Reads number	Total base(bp)	Read N50	Mean	Maximum
16,283,503	140,912,539,579	13,784	8,654	128,492

Table S2. Statistics of assembled *M. officinalis* assembly. Related to Table 1

	Canu assembly	Corrected assembly	Hi-C assembly
Contig number	12,562	10,294	11,562
Contig length (bp)	1,558,051,600	1,684,361,614	1,684,361,614
Contig N50 (bp)	207,504	302,541	222,069
Contig N90 (bp)	55,725	61,780	61,049
Contig max (bp)	1,499,188	4,084,461	2,700,203
GC content (%)	40.61	40.65	40.65

Table S3. Genome completeness assessment of *M. officinalis* genome by BUSCO.
Related to Table 1

Categories	Number	Percent (%)
Complete BUSCOs	1391	86.20
Complete and single-copy BUSCOs	1307	81.00
Complete and duplicated BUSCOs	84	5.20
Fragmented BUSCOs	75	4.6
Missing BUSCOs	148	9.2

Table S4. Completeness analysis of *M. officinalis* genome based on CEG database.
Related to Table 1

Number ¹ (458 of CEGs)	Percent ¹ (%)	Number ² (248 of CEGs)	Percent ² (%)
436	95.19	228	91.93

Note: **1.** Number and percentage of 458 core eukaryotic genes (CEGs) included in the genome assembly. **2.** Number and percentage of 248 highly conserved CEGs included in the genome assembly.

Table S5. Statistics of the genome assembly using Hi-C technology. Related to Fig. 1

Group	Sequence Number	Sequence Length (bp)
Lachesis Group0	992	144,591,505
Lachesis Group1	983	135,993,026
Lachesis Group2	841	119,381,364
Lachesis Group3	568	118,168,319
Lachesis Group4	701	108,152,317
Lachesis Group5	643	105,122,070
Lachesis Group6	560	95,926,108
Lachesis Group7	601	83,529,524
Lachesis Group8	575	83,604,872
Lachesis Group9	548	77,986,713
Lachesis Group10	546	77,939,050
Lachesis Group11	486	71,480,007
Lachesis Group12	375	70,748,615
Lachesis Group13	514	70,304,449
Lachesis Group14	529	69,875,511
Lachesis Group15	494	69,493,039
Lachesis Group16	603	68,551,582
Lachesis Group17	529	65,311,672
Lachesis Group18	382	42,497,259
Total Sequences Clustered (Ratio %)	11,470 (99.20)	1,678,657,002(99.66)
Total Sequences Ordered and Oriented	8,689(75.75)	1,531,087,320(91.21)

Table S6. The repetitive sequences of *M. officinalis*. Related to Table 1

Type	Number	Length	Rate (%)
ClassI	2,645,203	1,196,475,697	71.00
ClassI/DIRS	139,036	105,317,930	6.25
ClassI/LARD	1,056,904	402,713,815	23.90
ClassI/LINE	156,049	50,266,419	2.98
ClassI/LTR/Copia	593,974	355,691,881	21.11
ClassI/LTR/Gypsy	645,893	463,875,128	27.53
ClassI/LTR/Unknown	15,897	7,168,871	0.43
ClassI/PLE	6,899	3,164,692	0.19
ClassI/SINE	10,692	2,008,171	0.12
ClassI/TRIM	16,503	12,059,312	0.72
ClassI/Unknown	3,356	952,288	0.06
ClassII	199,143	95,899,299	5.69
ClassII/Crypton	19	1,181	0.00
ClassII/Helitron	26,147	7,612,587	0.45
ClassII/MITE	22,139	5,501,550	0.33
ClassII/Maverick	8,262	4,064,803	0.24
ClassII/TIR	138,169	77,885,592	4.62
ClassII/Unknown	4,407	1,452,070	0.09
PotentialHostGene	46,118	11,398,385	0.68
SSR	8,450	8,797,374	0.52
Unknown	722,215	211,549,783	12.55
Total	3,621,129	1,372,402,279	81.44

Table S7. Noncoding RNA prediction of *M. officinalis*. Related to Table 1

Type	Number	Family
rRNA	449	4
tRNA	575	24
miRNA	72	24

Table S8. Gene functional annotation of *M. officinalis* based on different database.
Related to Table 1

Database	Annotated number	Percentage (%)
GO	13,438	57.37
KEGG	8,253	35.23
KOG	13,845	59.11
TrEMBL	23,024	98.29
NR	23,035	98.34
ALL	23,050	98.40

Table S9. GO functional annotation in *M. officinalis*. Related to Fig. S2

Functional category	Term description	gene number
cellular component	extracellular region	393
cellular component	cell	5621
cellular component	nucleoid	10
cellular component	membrane	3206
cellular component	virion	3
cellular component	cell junction	230
cellular component	membrane-enclosed lumen	174
cellular component	macromolecular complex	1480
cellular component	organelle	4181
cellular component	extracellular region part	8
cellular component	organelle part	2030
cellular component	virion part	3
cellular component	membrane part	1419
cellular component	cell part	5675
cellular component	supramolecular complex	61
molecular function	transcription factor activity, protein binding	28
molecular function	nucleic acid binding transcription factor activity	156
molecular function	catalytic activity	7252
molecular function	signal transducer activity	121
molecular function	structural molecule activity	446
molecular function	transporter activity	869
molecular function	binding	6180
molecular function	electron carrier activity	267
molecular function	antioxidant activity	158
molecular function	metallochaperone activity	3
molecular function	protein tag	2
molecular function	translation regulator activity	1
molecular function	nutrient reservoir activity	55
molecular function	molecular transducer activity	66
molecular function	molecular function regulator	92
biological process	reproduction	685
biological process	cell killing	1
biological process	immune system process	165
biological process	metabolic process	7709
biological process	cellular process	7053
biological process	reproductive process	679
biological process	biological adhesion	34
biological process	signaling	502
biological process	multicellular organismal process	809

biological process	developmental process	1213
biological process	growth	239
biological process	locomotion	11
biological process	single-organism process	5381
biological process	biological phase	40
biological process	rhythmic process	22
biological process	response to stimulus	2538
biological process	localization	1946
biological process	multi-organism process	410
biological process	biological regulation	2173
biological process	cellular component organization or	
	biogenesis	1469
biological process	detoxification	205

Table S10. KEGG functional annotation in *M. officinalis*. Related to Fig. S2

pathway	Pathway id	Gene number
Carbon metabolism	ko01200	237
Biosynthesis of amino acids	ko01230	217
Starch and sucrose metabolism	ko00500	207
Phenylpropanoid biosynthesis	ko00940	190

Table S12 Gene family classification statistics. Related to Fig. 3

	Number of genes	Number of genes in orthogroups	Number of orthogroups containing species	Number of species-specific orthogroups	Number of genes in species-specific orthogroups
<i>A. thaliana</i>	27,449	18,388	12,533	2,482	6,715
<i>A. trichopoda</i>	16,839	15,516	12,770	191	602
<i>C. kanehirae</i>	26,531	23,777	14,908	218	630
<i>C. sinensis</i>	50,447	43,514	18,792	2,475	11,923
<i>G. biloba</i>	32,831	20,958	10,780	1,496	5,863
<i>L.chinense</i>	35,269	30,992	15,076	968	9,416
<i>L. cubeba</i>	30,805	26,113	16,149	475	1,364
<i>M. officinalis</i>	23,424	21,966	14,082	202	582
<i>N. colorata</i>	31,475	22,474	13,640	1,207	5,511
<i>O. sativa</i>	28,043	20,456	16,392	781	2,199
<i>P. americana</i>	22,440	19,573	14,562	131	283
<i>P. nigrum</i>	62,669	44,856	18,455	4,127	14,889
<i>P. trichocarpa</i>	31,453	29,858	16,365	798	2,879
<i>Z. mays</i>	34,104	29,557	18,035	1,517	5,198

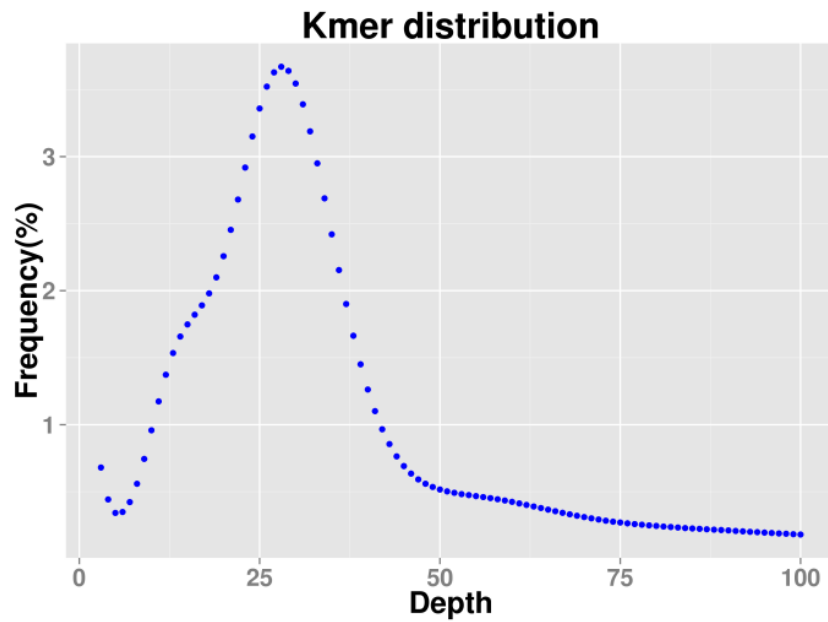


Fig. S1 K-mer frequency distribution curve of the *M. officinalis* genome. The horizontal axis represents the depth of coverage of the k-mer, and the vertical axis represents the frequency of the depth of the k-mer. The results showed that the average k-mer depth corresponding to the main peak was 28, with a total of 49,503,852,345 k-mers were used. The genome size was estimated based on the formula: Genome size = total k-mer number/average k-mer depth. Thus, we estimated *M. officinalis* genome size was approximately 1.76 Gb. Related to Fig. 2.

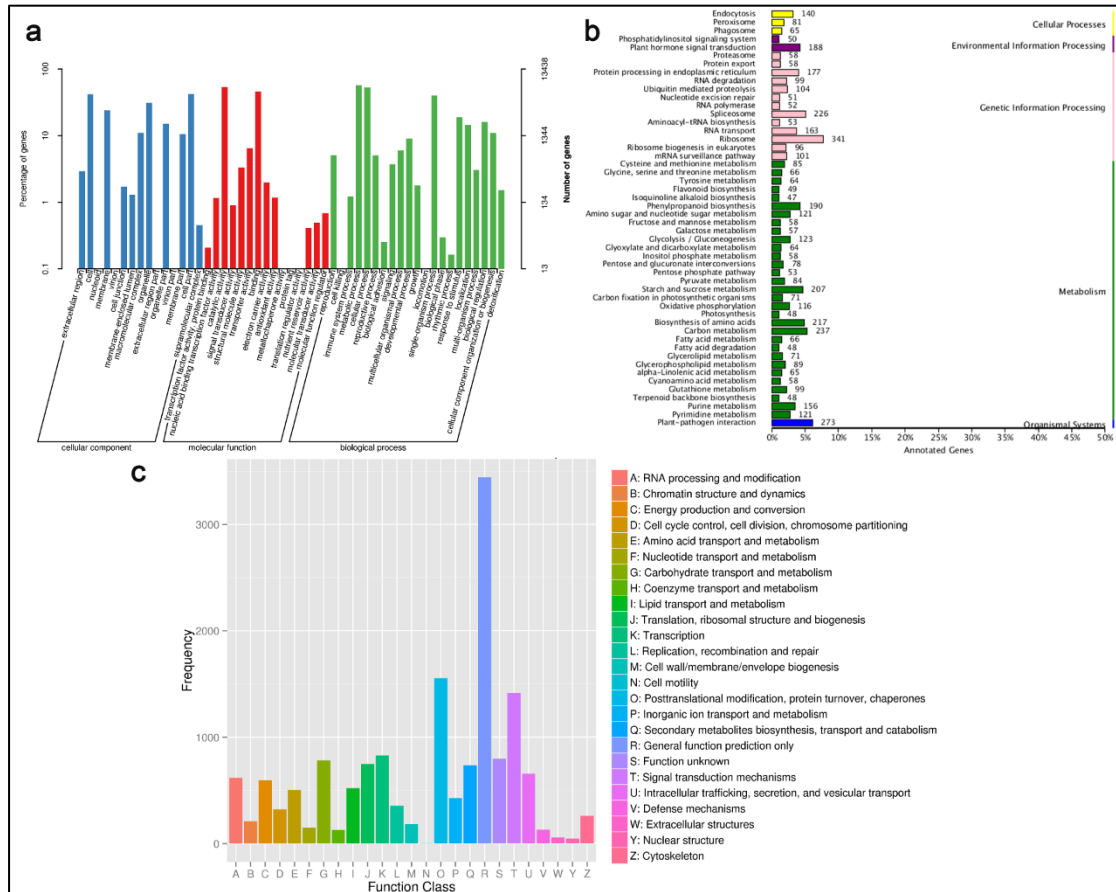


Fig. S2 Results of genomic enrichment analysis of *M. officinalis*. (a) Gene ontology categories of the annotated genes. The y-axis represents the count of genes identified in this study. The genes were annotated in three main categories: cellular component, molecular function and biological process (x-axis). (b) KEGG pathway of the annotated genes. The genes were annotated in many pathways. (c) KOG categories of the annotated genes. The KOG database can be divided into 26 categories by function. Related to Table S8.

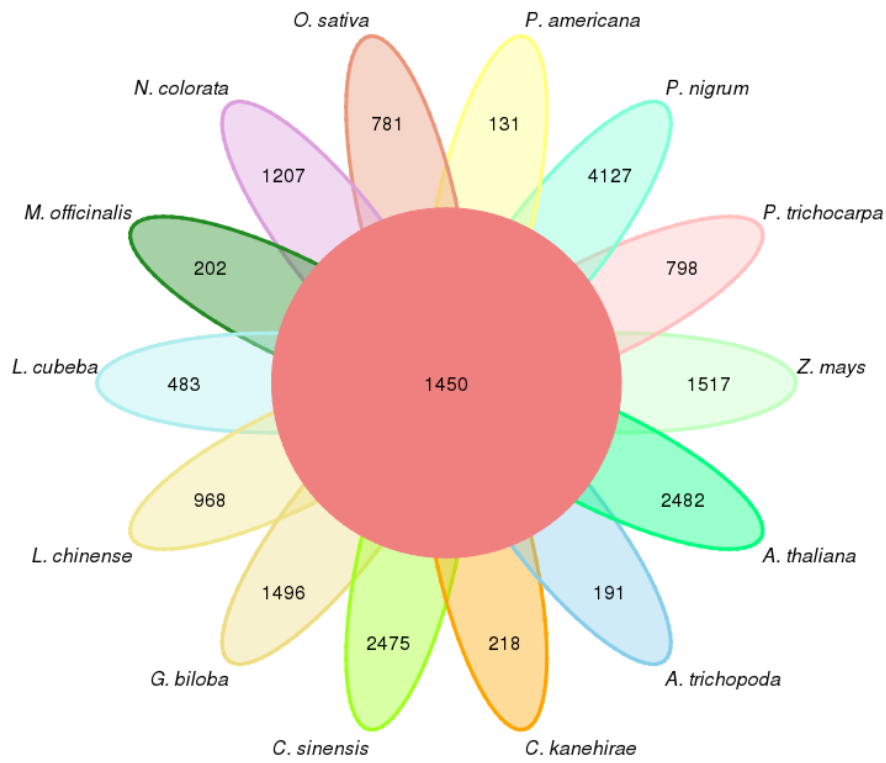


Fig. S3 Venn diagram of shared gene families in 14 species. The number of gene families is listed for each component. Related to Fig. 3.

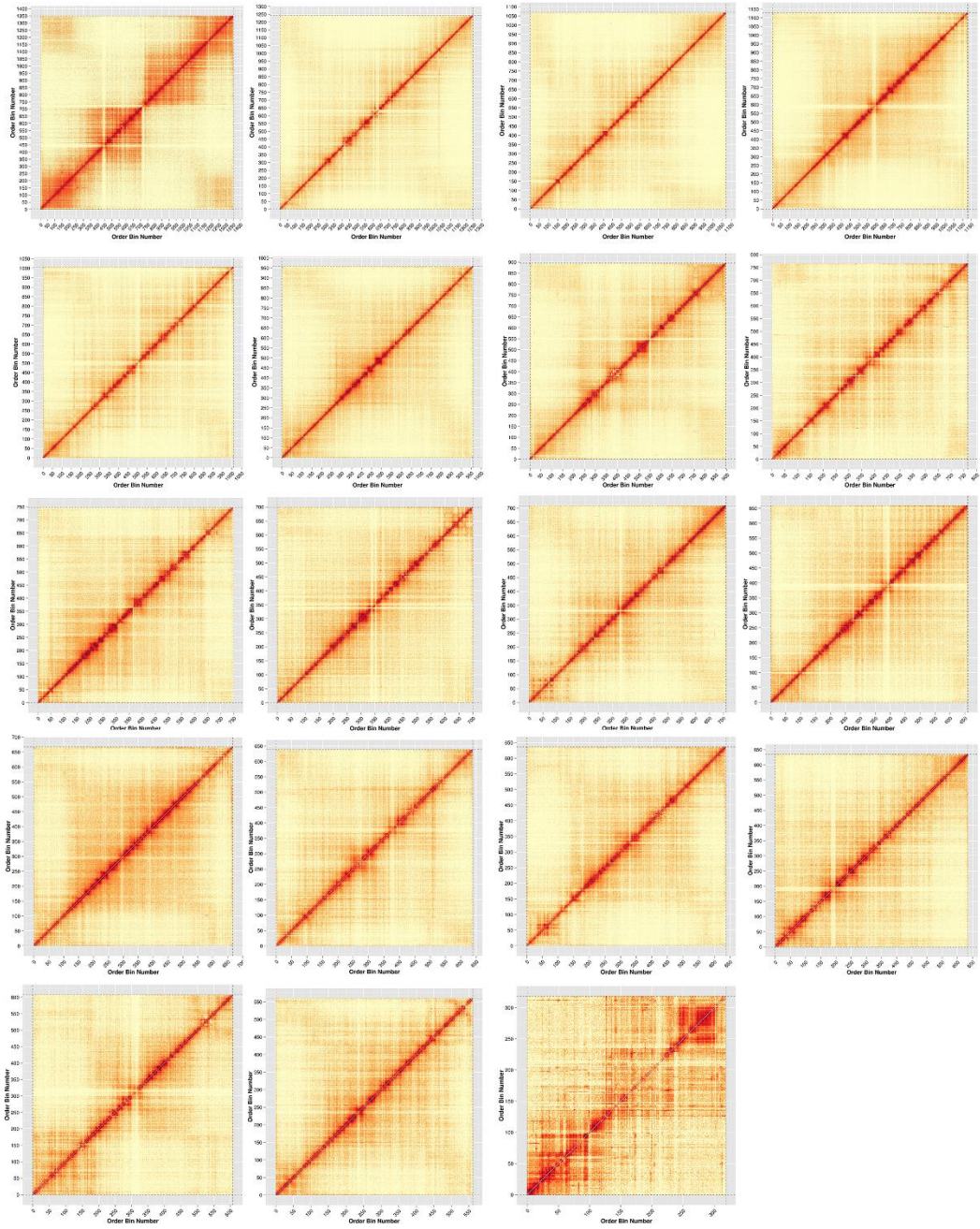


Fig. S4 Chromatin interactions in each chromosome of *M. officinalis*. The Hi-C data in *M. officinalis* were mapped to the *M. officinalis* genome. Related to Fig. 2.